

## Response to the Consultations on the White Paper on Online Harms

### ARTICLE 19

June 2019

1. ARTICLE 19 is an international human rights organisation which works around the world to protect and promote the right to freedom of expression and information (freedom of expression). With an international office in London and regional offices in Tunisia, Senegal, Kenya, Mexico, Brazil and Bangladesh, and other regional programmes and national offices, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, develops long-term strategies to address them and advocates for the implementation of the highest standards of freedom of expression nationally and globally. ARTICLE 19 has extensive expertise in the area of intermediary liability and regulation of online content. For example, we analysed the German Law on Enforcement of Illegal Content Online (so called NetzDG),<sup>1</sup> the French Law on Fake News,<sup>2</sup> the EU Code on Conduct on Countering Illegal Hate Speech Online<sup>3</sup> and responded to a series of EU consultations on notice and action procedures and how to tackle illegal content online.<sup>4</sup> We are therefore well-placed to comment on proposals to regulate tech companies for the purposes of preventing online ‘harms,’ as outlined in the White Paper on Online Harms (the White Paper).
2. Our response is structured into two parts. First, we set out general concerns over the overall approach of the White Paper to regulation. Second, we respond to questions set by the Government in the areas in which we have expertise.

#### General comments

3. At the outset, ARTICLE 19 notes that the Internet ecosystem has significantly evolved since intermediary liability laws were first adopted in the late 1990s - early 2000. Major social media companies such as Facebook, YouTube or Twitter have become fundamental to how people communicate and hold exceptional influence over individuals’ exercise of their right to freedom of expression online. Importantly, the roles played by these companies have evolved in recent years and now range from hosting (a role characterised by the absence of editorial intervention on content) to actively promoting selected content (making selected content more visible through human or algorithmic means) or even actively producing content. In the UK (as well as in other countries), these companies benefit from a regime of conditional immunity from liability for hosting illegal content. However, the power of some of these companies in our society and their dominance in several markets raise legitimate concerns about their accountability to the wider public. This is an important debate to be had and states have a role to play in

---

<sup>1</sup> ARTICLE 19, [Germany: Act to Improve Enforcement of the Law on Social Networks undermines free expression](#), 1 September 2017.

<sup>2</sup> ARTICLE 19, [France: Proposed law against the manipulation of information violates international standards](#), 23 July 2018.

<sup>3</sup> ARTICLE 19, [EU: European Commission’s Code of Conduct for Countering Illegal Hate Speech Online and the Framework Decision](#), 20 August 2016.

<sup>4</sup> ARTICLE 19, [EU: Commission approach to tackling illegal content online fails to protect free expression](#), 10 July 2018.

redressing the power imbalance between tech giants, in particular some social media platforms, and other actors, including users, developers and other smaller competitors.

4. ARTICLE 19 believes that these developments require a careful examination of the requirements of freedom of expression when it comes to digital platforms. To this end, states must identify what is the necessary and least restrictive method to achieve effective protection of each of the objectives traditionally assigned to regulation (such as pluralism and diversity of freedom of expression), taking into account the evolution and roles of digital platforms in promoting and protecting human rights, including freedom of expression online.
5. However, ARTICLE 19 is concerned that instead of holistically considering what framework might be appropriate to regulate digital platforms, the proposal in the White Paper is unduly focusing on online content regulation. This is at the expense of solutions to excessive market concentration and abuse of dominance of these companies that could be found in competition, consumer protection and data protection law, among others. Although the UK Government's efforts at regulating digital platforms may be filled with good intentions, ARTICLE 19 has significant concerns with the way in which the White Paper is framed and its scope. While some effort seems to have gone into giving this regulatory endeavour some conceptual underpinning with a new 'duty of care,' we believe that it is sorely lacking in clarity and could pose a significant threat to freedom of expression. It is also highly unclear that the proposed solutions would ultimately be effective. In particular, we raise the following key concerns.

#### ***Lack of an evidence-based approach***

6. ARTICLE 19 believes that the UK Government should pause before adopting of legislation that could prove damaging for freedom of expression without solving the root causes of the 'harms' identified in the White Paper. In our view, the Government should first identify more clearly the 'prevalence' and impact of the 'harms' at issue based on independent research and comprehensive scientific data<sup>5</sup> – these are currently missing in the proposal. This would enable the design of more effective and evidence-based policy solutions. In our view, this should include the provision of adequate resources to address the underlying offline behaviour at issue through media literacy programmes and law enforcement of unlawful conduct where appropriate. Government should also carry out a proper impact assessment of any proposed measures and identify any unintended or counter-productive effects, as well as any anti-competitive outcomes. We provide a more detailed analysis of the White Paper proposals further below.

#### ***Adversarial framing***

7. ARTICLE 19 regrets the adversarial tone of the White Paper and its tendentious approach to the issues at stake. The entire framing of the White Paper is built around negative emotions of fear and a state of helplessness that demand strong government action to provide safety and protection from nebulous 'harms'. For instance, the White Paper places significant emphasis on child abuse images and terrorism from the very beginning as justification for the need for regulation of 'online harms'. Whilst we recognise that these are serious issues, they are also the kind of well-known populist tropes often favoured by governments in order to whip up public support for otherwise-disproportionate measures. Further, the White Paper repeatedly uses emotionally charged language, describing content as 'appalling' or 'horrifying,' apparently to give a sense of urgency to action being taken by the Government. More often than not,

---

<sup>5</sup> Victoria Baines, [On Online Harms and Folk Devils: Careful Now](#), 24 June 2019.

however, the White Paper fails to establish a strong correlation between any ‘harm’ done and the need for the wide-ranging regulation envisaged by the Government.<sup>6</sup> Nor does it explain how regulation would solve problems that originate in the behaviour of individuals in the offline world (e.g. terrorism or child abuse). We note that the UN Special Rapporteur on freedom of expression recently warned against this approach, stating “a ‘rhetoric of danger’ is exactly the kind of rhetoric adopted in authoritarian environments to restrict legitimate debate, and we in the democratic world risk giving cover to that.”<sup>7</sup> In our experience, this kind of crude politicking is not conducive to good policy-making.

8. ARTICLE 19 also notes that the tone of the White Paper proposal is in marked contrast with the far more constructive tone and positive framing of the recent **interim report on social media regulation** of the French Government. The French model places transparency and accountability by design at the heart of its proposals.<sup>8</sup> In our view, should any draft legislation be laid before Parliament in this area, it should follow the same approach and put the protection of human rights at its core. For instance, it could provide incentives for companies to respect the UN Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework (the Guiding Principles) as a starting point for articulating the role of the private sector in protecting human rights on the Internet.<sup>9</sup> In practice, this would mean transparency obligations, including in relation to companies’ community standards and internal processes. It would also involve human rights impact assessments in relation to companies’ operations and obligations to put in place takedown processes and internal complaints mechanisms that are consistent with due process standards.<sup>10</sup>

#### ***Overbroad scope of the White Paper***

9. **Subject matter – legally protected speech:** ARTICLE 19 finds it extremely worrying that the proposed regulatory framework is not confined to illegal content.<sup>11</sup> The White Paper proposes that the new regulation would cover ‘legal but harmful’ content – i.e. content that is not prohibited under domestic legislation. We are concerned that the White Paper does not seek to define these ‘harms’ but instead leaves it to the regulator to define. Moreover, the White Paper suggests that the list of ‘harms’ in scope should be open-ended so as to enable a regulator to address any emerging ‘harms.’<sup>12</sup>
10. In ARTICLE 19’s view, the Government’s approach to online ‘harms’ is deeply problematic. We note that under international human rights law, restrictions on freedom of expression are only permissible if they are (1) provided by law, (2) pursue a legitimate aim, expressly enumerated in Article 19(3) of the International Covenant on Civil and Political Rights or Article 10(2) of the European Convention on Human Rights and (3) are necessary and proportionate to the aim sought. We believe that the lack of definition of ‘legal but

<sup>6</sup> *Ibid.*

<sup>7</sup> See Peter Pomerantsev, [How \(Not\) to Regulate the Internet, The American Interest](#), 20 June 2019.

<sup>8</sup> [Creating a French framework to make social media platforms more accountable: Acting in France with a European vision](#), Mission report “Regulation of social networks – Facebook experiment,” Submitted to the French Secretary of State for Digital Affairs, May 2019.

<sup>9</sup> [Guiding Principles on Business and Human Rights: Implementing the UN ‘Protect, Respect and Remedy’ Framework](#), developed by the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations Endnotes 44 and Other Business Enterprises, John Ruggie, 7 April 2008, A/HRC/8/5A/HRC/17/31. The Human Rights Council endorsed the Guiding Principles in its resolution 17/4 of 16 June 2011.

<sup>10</sup> See ARTICLE 19, [Sidestepping Rights: Regulating Speech by Contract](#), 2018.

<sup>11</sup> *C.f.*, for example, the German NetzDG.

<sup>12</sup> The White Paper, para. 2.2.

harmful’ content in the legislation implementing the White Paper proposal would fail to meet the first limb of this test. The White Paper refers extensively to various forms of ‘harm’ which are ill-defined (e.g. ‘harms with a less clear definition’) but also, in many respects, highly speculative. There is, for example, no agreement on what constitutes ‘violent content,’ still less that it causes ‘harm’. The same is true of disinformation, for instance.<sup>13</sup> The White Paper, however, frequently elides content which might be seen as undesirable, on the one hand, with the idea that such content is ‘harmful.’ It should be obvious to any reasonable person that it is not necessary for something to be ‘harmful’ in order to be undesirable, and equally obvious that the law permits people to do many things that might be viewed by others as undesirable. The White Paper, however, makes no attempt to distinguish between these two categories.

11. More generally, it is concerning that the Government considers it appropriate to seek to regulate online content on the basis of a concept of ‘harm’ that it is either unwilling or unable to define. Instead, the White Paper’s repeated reference to ‘harmful content’ seems to cast it as a sweepingly-broad category that covers ‘bad things that no sensible or respectable person would want to see,’ i.e. an ill-defined and subjective concept that is plainly untenable as the basis for sensible legislation and incompatible with the legality principle under international human rights law.
12. ARTICLE 19 believes that, instead of trying to apply broadcast-type regulation to both illegal and ‘legal but harmful’ online content, the Government should seek to clarify intermediary liability rules by laying down clear procedures for the removal of *illegal* content that are in line with international standards on freedom of expression and due process.<sup>14</sup> This should be coupled with more resources for the enforcement of existing laws against criminal conduct. If the Government nonetheless pursues its current approach, i.e. involving regulatory oversight, we believe that at a minimum, it should remove ‘legal but harmful’ content from the scope of the regulation.<sup>15</sup> In this kind of scenario, the regulator should not be involved in the determination of the legality of content, but instead focus on transparency obligations and reviewing internal company processes on content moderation. If the Government does not heed our advice, we believe that at the very least it should make clear that companies’ obligations in relation to ‘legal but harmful’ content do not include content removal or other forms of restriction on content that would be meted out with sanctions since this would encourage over-removal. Instead, the Government should focus on greater transparency and accountability mechanisms in the application of companies’ terms of service/community standards.
13. **Companies in scope:** ARTICLE 19 notes that the White Paper is currently intended to cover “companies that allow users to share or discover user-generated content or interact with each other online.” This would not be limited to social media companies but would also include file-hosting sites, public discussion forums, messaging services and search engines. It would also cover the comment sections on websites of media outlets or individuals.

<sup>13</sup> See for instance King’s College London Centre for the Study of Media, Communication and Power, [Submission to the Culture Media and Sport Select Committee’s Inquiry into Fake News](#), 2017.

<sup>14</sup> See UN Special Rapporteur on freedom of expression, A/HRC/38/35, 6 April 2018 at paras. 66-68 and, for instance, [the Manila Principles on Intermediary Liability](#).

<sup>15</sup> In this respect, we note that the equivalent German legislation (NetzDG) in this area is limited to illegal content.

14. In our view, the scope of the White Paper is overly broad. In particular, it is likely to impose sweeping demands on companies which would quite simply not be in a position to comply with such a regulatory burden. It is also likely to raise significant issues for the protection of the right to privacy, particularly as regards services that users rightly consider to be private such as messaging apps. We note the Secretary of State Jeremy Wright MP has given assurances that the press would be excluded, providing that they are members of one of the current press regulators in the UK (IMPRESS and IPSO).<sup>16</sup> The potential exceptions for media outlets that are members of two UK regulators are also a matter of concern, in particular in light of existing issues with press regulation.<sup>17</sup> This would also mean that individuals that allow third-party interactions on their websites would be subject to stricter regulation than major new media outlets. We therefore urge the UK Government to re-consider the companies in scope of any new regulatory framework and limit it to only dominant social media platforms, i.e. the biggest players in terms of user-base, content reach and turnover and those providing communication services as their main service. We elaborate on some of these points further below.

***Lack of clarity of the concept of ‘duty of care’***

15. ARTICLE 19 regrets the lack of conceptual clarity in the proposed approach in the White Paper. It borrows concepts from tortious liability (‘the duty of care’). We find that this concept raises serious concerns when applied to online content moderation.<sup>18</sup> We note that in general, duties of care arise when harm is the reasonable foreseeable result of a defendant’s conduct, there is a relationship of proximity between the defendant and the claimant and it is fair, just and reasonable to impose liability. Breach of the duty leads to liability in the common law sense. However, the White Paper does not envision that its undefined ‘duty of care’ would apply to digital companies in this fashion. For instance, there is no suggestion in the White Paper that users could bring a claim in negligence against the companies who fail to comply with their ‘duty of care,’ for instance by failing to remove content. In other words, the ‘duty of care’ appears to be no more than a confusing label when the Government is in fact proposing a traditional regulatory framework akin to broadcast regulation.

16. In any event, a key problem with the White Paper proposal is that it fails to define key terms, including ‘the duty of care’ and ‘risk of harm;’ these would be left to the regulator to determine. In ARTICLE 19’s experience, a ‘duty of care’ is usually synonymous with upload filters and removals within unduly short time frames for various types of content, such as ‘terrorism’ or ‘hate speech’ online. In our view, building an entire regulatory framework on the basis of terms, which are unacceptably broad and hard to define, is a recipe for failure. As such, we believe that the Government should abandon the ‘duty of care’ approach and focus instead on transparency obligations coupled with obligations to set up internal complaints mechanisms. It should also ensure that users have access to independent redress. If, however, the Government retains the term ‘duty of care’ in the legislation, it should at least define it and consider the ten-point rule of law test for a social media duty of care set out by leading experts in the field.<sup>19</sup>

<sup>16</sup> See Jeremy Wright’s [Letter to the Society of Editors](#), April 2019.

<sup>17</sup> ARTICLE 19, [UK: Decision to halt Leveson threatens effective self-regulation of the press](#), 9 March 2018.

<sup>18</sup> See Cyberleagle, [Take Care with that Social Media Duty of Care](#), 19 October 2018.

<sup>19</sup> See Cyberleagle, [A Ten Point Rule of Law Test for a Social Media Duty of Care](#), 16 March 2019; [The Rule of Law and the Online Harms White Paper](#), 5 May 2019; and [Speech is not a tripping hazard: response to the Online Harms White Paper](#), 28 June 2019.



17. Finally, ARTICLE 19 notes that the White Paper does not clearly articulate how the new regulatory framework would fit with existing principles of intermediary liability. It purports to be consistent with the EU Directive on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (the E-Commerce Directive), when some of its stated aims are plainly at odds with the wording of the E-Commerce Directive, particularly when it comes to the prohibition on general monitoring. It is entirely unclear, for example, how companies are expected to deal with ‘harmful but legal’ content short of general monitoring. The White Paper also fails to clarify the extent to which the intermediary liability framework derived from the E-Commerce Directive would remain applicable, e.g. whether companies who comply with the so-called ‘duty of care’ could still be found liable for failure to remove specific items of content upon notice. In short, while the White Paper might look superficially thoughtful and comprehensive, it is in fact piecemeal, lacking in detail and conceptually muddled.

### ***Disproportionate sanctions***

18. ARTICLE 19 further notes that the sanctions proposed in the White Paper, which range from severe fines to website blocking of non-compliant sites and senior management liability, are far-reaching and unlikely to meet the proportionality test under international human rights law (as outlined earlier).<sup>20</sup> In particular, we note that a regulator’s power to fine social media companies could only be compatible with the right to freedom of expression if it relates to clear and narrowly defined obligations and any fines imposed are proportionate to the gravity of the conduct at issue. This is clearly not the case at present.
19. Rather than embracing a highly punitive approach, we urge the Government to explore alternatives that would reward companies for demonstrating higher standards of conduct. This could include kite-marking or grading that would enable the public to recognize companies that abide by higher standards of conduct. We would also encourage the government to consider independent multi-stakeholder models, such as Social Media Councils,<sup>21</sup> that would allow public debate and independent oversight of key issues in content moderation without unduly harsh sanctions.

## **Response to key questions in the White Paper**

### **Question 1 - Transparency reporting**

20. **ARTICLE 19 appreciates the Government’s stated commitment to** “developing a culture of transparency, trust and accountability” in the White Paper. Under the Government’s proposals, the regulator would have the power to require annual transparency reports from companies in scope. The transparency report would highlight the “prevalence of harmful content” on the platforms and set out what counter measures the companies are taking to address them. The regulator would also have powers to require additional information, including about the impact of algorithms in selecting content for users and to ensure that companies proactively report on both “emerging and known harms.”

<sup>20</sup> *Op.cit.* See also the 2018 thematic report of the UN Special Rapporteur on freedom of expression, who calls on States to “refrain from imposing disproportionate sanctions, such as heavy fines or imprisonment, on Internet intermediaries;” A/HRC/38/35, para 66.

<sup>21</sup> See, e.g. ARTICLE 19, [Self-regulation and ‘hate speech’ on social media platforms](#), 2018.

21. ARTICLE 19 notes that the government’s approach to transparency reporting is unduly limited by its focus on ‘online harms.’ In our view, transparency should be a basic requirement that pervades everything that companies do. In particular, it should apply to:

- *Distribution of content*: digital companies should explain to the public how their algorithms are used to present, rank, promote or demote content. Content that is promoted should be clearly marked as such, whether the content is promoted by the company or by a third-party for remuneration.
- *Companies’ terms of service and community standards*: companies should publish community standards/terms of service that are easy to understand and give “case-law” examples of how they are applied. As suggested by the French Government interim report (see above), they should publish information about the methods and internal processes for the elaboration of community rules.
- *Human and technological resources used to ensure compliance*: companies should include detailed information about trusted flagger schemes, including who is on the roster of trusted flaggers, how they have been selected and any ‘privileges’ attached to that status. They should also publish information about the way in which their algorithms operate to detect illegal or allegedly ‘harmful’ content under their community standards. In particular, this should include information about rates of false negatives/false positives and indicators, if any, to assess content that is likely to become viral, e.g. by reference to exposure to a wider audience.
- *Decision-making*: companies should notify their decisions to affected parties and give sufficiently detailed reasons for the actions they take against particular content or accounts.<sup>22</sup> They should also provide clear information about any internal complaints mechanisms.
- *Transparency reports*: companies should publish detailed information consistent with the Santa Clara Principles that have been developed by experts in the field.<sup>23</sup> We note that it is particularly important not to limit statistical information to removal of content but also include data about the number of appeals processed and their outcome. Transparency reporting should also distinguish between content flagged by third-parties (including whether they are public bodies or private entities), trusted flaggers (whether public bodies or private entities) or algorithms.

22. More generally, we note that any transparency reporting requirements should aim to provide far more qualitative analysis of content moderation decisions. It is vital that the metric of success in relation to the fulfilment of any ‘duty of care’ is not tied to content removal rates as it encourages over-removal.<sup>24</sup> Equally, transparency reporting should not be limited to information submitted by companies but should include information submitted by relevant government agencies, such as the number of takedown requests issued by the Counter-Terrorism Internet Referrals Unit (‘CTIRU) or the Police Intellectual Property Unit (PIPU).

#### **Recommendations:**

- Transparency should be a basic requirement that pervades everything that companies do, including the distribution of content, their terms of service and community standards,

<sup>22</sup> *Ibid.*

<sup>23</sup> [The Santa Clara Principles on Transparency and Accountability of Content Moderation Practices](#), 2017.

<sup>24</sup> See, e.g., Heidi Tworek, [An Analysis of Germany’s NetzDG Law](#), 15 April 2019.

human and technological resources used to ensure compliance, decision-making and transparency reports;

- Transparency reporting and the metric of success for compliance with any ‘duty of care’ should not be limited to takedown rates but aimed at a qualitative analysis of content moderation decisions.

## Question 2 - User redress and ‘super complaints’

23. ARTICLE 19 believes that governments have a duty to ensure users’ right to an effective remedy to deal with infringements of their rights, including through a right of appeal to the courts or other independent body. We have also long advocated for companies to provide internal complaints mechanisms, including appeals against wrongful removals of content, account suspensions or strikes against user accounts.<sup>25</sup> We therefore welcome the inclusion of a section on ‘user redress’ in the White Paper that contemplates both internal and external appeals processes.

24. Nonetheless, ARTICLE 19 remains concerned that the UK Government’s proposal on user redress, outlined in the White Paper, fails to pay due regard to freedom of expression and is unduly narrow in scope. In particular, we highlight the following issues:

- *Lack of focus on protecting freedom of expression*: we are concerned that in keeping with the government’s focus on ‘online harms,’ the White Paper makes no mention of the possibility of a remedy for wrongful removal of content. This, in our view, is a striking omission given the important role that platforms play in promoting freedom of expression. Similarly, the ‘super complaints’ process does not mention whether it would be available for instances of ‘serial’ wrongful removal of content, i.e. where content is repeatedly removed in error or out of a superabundance of caution. In our view, it is vital for the government to be explicit that a regulator would be tasked with protecting freedom of expression and ensuring that content is not unduly removed. These principles should guide both the design and any oversight of internal complaints mechanisms or super complaints processes. In practice, this means strong procedural safeguards, including evidentiary standards.
- *Unduly restrictive approach to user redress*: we believe that the Government’s approach to user redress appears to be too limited. The White Paper states that it does not envisage a role for the regulator itself in determining disputes between individuals and companies. While this may be appropriate, any legislation in this area should provide for a broad right of individual redress, subject to traditional admissibility criteria and limitations on abuse of process. This is true both in relation to internal mechanisms and independent review processes.

Further, in our view, scale is not a sufficient reason to unduly limit the right of redress. We note, for instance, that appeals are widely available under data protection law despite the fact that potentially millions of users may complain about data breaches. There is no reason in principle why a similar redress mechanism should be denied to individuals complaining about wrongful removal of content. Equally, since millions of individuals are given the ability to report content without the content provider being given an opportunity to respond prior to removal, basic fairness demands that the right of redress should be widely available and not restricted to certain categories of

---

<sup>25</sup> See Sidestepping Rights: Regulating Speech by Contract, *op.cit.*



content - though it may be subject to traditional admissibility criteria. Insofar as government or companies may be minded to limit redress mechanisms available for particularly egregious types of content, such as child abuse images or terrorist content, we note for instance that the Internet Watch Foundation has a content assessment appeal process.<sup>26</sup> Similarly, the recent Christchurch Call against Terrorist and Violent Extremist Content makes reference to complaint mechanisms.<sup>27</sup>

In short, ARTICLE 19 believes that internal complaints mechanisms and independent reviews should be available across the board in relation to all types of content. In practice, this means that an independent body should be able to adjudicate individual complaints about the extent to which companies have applied their terms of service in a way that is consistent with international standards on human rights and/or equivalent constitutional protection of those rights. For example, ARTICLE 19 is exploring the model of Social Media Councils, which would also include the possibility of individual redress and how this could work.<sup>28</sup> In principle, a Social Media Council would only have the power to impose *non-pecuniary* remedies, including a right of reply, an apology or re-instatement of content.<sup>29</sup> Finally, insofar as companies take down content to comply with court orders or orders made by other public bodies, individuals should have a right to challenge those orders in court.

25. Finally, ARTICLE 19 recommends that the government should clarify the extent to which the remedies envisaged in the White Paper would sit in relation to the E-Commerce Directive.

**Recommendations:**

- Users should have access to internal complaints mechanisms for wrongful removal of content;
- Users should have a right to an effective remedy, including access to independent redress mechanisms for wrongful removal of content.

**Question 4 - Parliamentary oversight**

26. As noted above, ARTICLE 19 believes that the Government should not seek to regulate ‘harmful’ content through Codes of Practice. Instead, it should focus on laying down clear procedures for the removal of illegal content online, consistent with due process safeguards and the protection of freedom of expression.<sup>30</sup> If the Government believes that particular ‘harms’ it has identified are sufficiently serious that companies should be *required* to take that type of content down, then Parliament should legislate to prohibit it. We do not think that it is good enough for the Government to shirk responsibility for defining ‘harmful’ content, passing on that responsibility onto the regulator. In any event, it seems highly doubtful that any institution – whether a regulator or Parliament itself - would be able to define some of the harms mentioned in the White Paper, such as ‘false information,’ ‘extremism’ or ‘bullying’ with a degree of precision that would be sufficient to comply with the legality principle. Yet, companies would be expected to comply or face the prospect of fines for systemic failures. Users’ expression is therefore likely to be

<sup>26</sup> Internet Watch Foundation, [Content assessment appeal process](#).

<sup>27</sup> [Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online](#).

<sup>28</sup> ARTICLE 19 and Stanford Global Digital Policy Incubator, [Social Media Councils: from Concept to Reality](#), 2019:

<sup>29</sup> See ARTICLE 19, [Social Media Councils: Consultation Paper](#), June 2019.

<sup>30</sup> See e.g. Manila Principles, *op.cit.* and UN Special Rapporteur on freedom of expression, *op. cit.*

censored on a massive scale, since perfectly legitimate content will inevitably get caught. Recent examples have already emerged in relation to YouTube's new policy on 'hateful and racist' videos: the work of journalists, researchers, professors has been removed at the same time as the 'hateful' content they were commenting on.<sup>31</sup>

#### Recommendations:

- If the government believes that particular 'harms' it has identified are sufficiently serious that companies should be *required* to take that type of content down, then Parliament should legislate to prohibit it.
- Legislation in this area could clarify content takedown procedures in line with international standards on freedom of expression and due process safeguards.

#### Questions 5, 6 and 7 - Companies in scope

27. ARTICLE 19 reiterates that the scope of the White Paper is both overly broad and impractical. If the main concerns around online content arise out of the practices of dominant social media companies, the scope of the White Paper should be limited to those companies.
28. If the government retains the current companies in scope, however, ARTICLE 19 agrees that a tiered approach to companies' obligations would be appropriate. This is also the approach adopted in the recent French report (see above). In our view, any tiered obligations should take into account the turnover, size and market power of the companies at issue among others. Smaller platforms or not-for-profit organisations such as Wikipedia can generally not sustain the same regulatory burden as big, for profit, organisations such as Facebook.
29. ARTICLE 19 believes that private communications channels and forums should be out of scope. They should not be subject to any particular obligations to monitor content or otherwise imposed by a regulator. ARTICLE 19 notes that under European human rights law, even ostensibly 'public' activities can fall within the scope of the right to privacy.<sup>32</sup> In our view, whether or not communications are private cannot be answered simply by reference to broad categories or number of users of a particular channel or messaging group. Rather, they are more likely to be a question of fact and degree, to be determined by reference to a number of factors. For instance, nobody would seriously suggest that an email containing an internal memo is not private, simply because it has been sent to hundreds of employees at a company. Equally, WhatsApp groups expect their communications to remain private, though having a large group of participants makes it more likely that leaks will occur. In practice, factors that are likely to be relevant include the nature of the information at issue, the context in which the information was shared, the means of communications used, and more generally any reasonable expectation of privacy. The latter would depend on factors such as whether access to the platform is restricted (password-protected), whether users rely on encrypted communications or other means of communication that have strong security protection. In any event, insofar as the government may recognise certain categories as 'private channels and forums', they should be out of scope.

---

<sup>31</sup> See LA Times, [YouTube's purge of white supremacist videos also hits anti-racism channels](#), 06 June 2019.

<sup>32</sup> See e.g. *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, [GC], no. [931/13](#), 27 June 2017

**Recommendations:**

- Only dominant social media platforms should be in scope;
- Private communications channels and forums should be entirely out of scope and not subject to any particular obligations to monitor content or otherwise imposed by a regulator.

**Question 8 – Proportionality**

30. ARTICLE 19 believes that the Government's approach in the White Paper is fundamentally flawed and that the regulation of 'online harms' should be abandoned. In any event, any regulator will be required to comply with the Human Rights Act, so the principles of necessity and proportionality ought to pervade everything it does. If – despite our warnings - legislation is thought necessary, it should at the very least be grounded in the protection of freedom of expression, in the same way that the protection of privacy ought to have been at the core of the Investigatory Powers Act 2016.
31. Additionally, we believe that it would be very important for any regulator to be tasked with protecting freedom of expression as part of its mission. A regulator should have a duty to consider and protect freedom of expression every time it is considering adopting measures that interfere with it. There should be clear provisions to that effect in legislation. In comparison, For instance, we note that a Private Members' Bill to regulate hate speech online in France (currently going through the National Assembly) would vest powers in the Conseil Supérieur de l'Audiovisuel (equivalent of Ofcom) to examine both whether communication service providers are failing to remove enough illegal content or are excessively removing it.<sup>33</sup> The French Bill also envisages the creation of a new summary offence of malicious notification to deter users from knowingly making false representations as to the legality of content. Whilst this approach is not without its problems, not least the difficulty in proving malice, it provides some protection for freedom of expression by penalising bad faith notices. Similar provisions can be found in the Digital Millennium Copyright Act (DMCA) in the US.<sup>34</sup>
32. Finally, we note that it would be highly improper – and inconsistent with international standards on freedom of expression - for companies to be penalised for failing to remove lawful speech.

**Recommendations:**

- If – despite our warnings - legislation is thought necessary, it should at the very least contain clear provisions to protect freedom of expression, including:
  - an overarching provision stressing the importance of protecting freedom of expression, including expression that may shock, offend or disturb;
  - a provision making clear that the mission of any regulator in this area is to protect human rights, including freedom of expression;
  - a provision requiring any regulator to audit content removal decisions and consider the extent to which companies over-remove content, whether upon request or of their own accord;
  - a provision making clear that companies should not be penalised for failing to remove lawful content.

<sup>33</sup> See Article 4 of the Draft [Bill on Countering Hate Speech Online](#) (version of 30 June 2019).

<sup>34</sup> See 17 USC 512(f).

### Question 10 - The regulator

33. ARTICLE 19 believes that the Government's approach to online content regulation is misguided. We do not support the setting up a new body or broadening the powers of existing ones for that purpose. We note, however, that a recurring problem with the regulation of the activities of certain actors is that several sets of rules apply at any one time, yet the relevant bodies tasked with the enforcement of those rules usually fail to coordinate. For instance, the regulation of the activities and behaviour of 'platforms' or 'information society providers' is relevant to the Competition and Markets Authority, and the Information Commissioner's Office, as well as Ofcom in respect of some electronic communications services. As a starting point, these regulators should be better coordinated. Equally, we note that consumer law is plainly relevant to companies' terms of service, yet it is unclear what role the Which organisation is playing in current debates about 'platform regulation'.<sup>35</sup>
34. Finally, we reiterate that bodies tasked with combatting particular categories of content (e.g. 'hate speech') or protecting a particular right (e.g. data protection), either tend to lack expertise in free expression law or inevitably err on the side of the most restrictive interpretation of that right.<sup>36</sup> More generally, bodies that traditionally apply an economic approach to regulation are not usually well equipped to carry the kind of balancing exercise inherent in human rights law.

#### Recommendation:

- Existing regulators should be better coordinated;
- Relevant rules should not be applied in silos.

### Questions 12 and 13 – Sanctions

35. ARTICLE 19 believes that the sanctions envisioned in the White Paper, which include powers to i) disrupt business activities, ii) undertake ISP blocking, or iii) implement a regime for senior management liability, would amount to a disproportionate interference with freedom of expression.
36. Disrupting business activities is a severe restriction on companies' right to freedom of expression and right to property. It is hard to see how it could possibly be justified in the absence of clear definitions of 'harm' or 'risk of harm.' The same is true of website blocking, which we have long criticised.<sup>37</sup> Equally, the UN Special Rapporteur on freedom of expression has expressed serious concerns about potentially similar senior management liability provisions in the recent Australian Criminal Code Amendment (Sharing of Abhorrent Violent Material) Law 2019 in the aftermath of the Christchurch terrorist attacks.<sup>38</sup>

<sup>35</sup> See, for instance, Euractiv, [Consumer watchdogs want fines against Twitter and Facebook for breaking EU rules](#), 15 February 2018.

<sup>36</sup> See, *mutatis mutandis*, Daphne Keller, [Free Expression Gaps in the General Data Protection Regulation](#), 6 December 2015.

<sup>37</sup> ARTICLE 19, [Freedom of Expression Unfiltered: How blocking and filtering affect free speech](#), 2016.

<sup>38</sup> UN Special Rapporteur on freedom of expression, [Comment on Criminal Code Amendment \(Sharing of Abhorrent Violent Material\) Law 2019](#), 4 April 2019.

37. More generally, it is unclear why such sanctions are strictly necessary. ISP/app blocking could be highly detrimental to Internet users who would be denied access to a range of services, some of which might be perfectly legal. Just because a platform may not be taking down certain categories of content (e.g. grossly offensive communications, which are currently criminalised in the UK) does not necessarily mean that a platform or service may not otherwise contain perfectly lawful information. Moreover, blocking tends to be a blunt instrument, which can be relatively easily circumvented. Its effectiveness is therefore highly questionable.
38. ARTICLE 19 cautions against any requirement imposed on companies based outside the UK to appoint a nominated representative in the UK or EEA. Nominated representatives should not be used to get around legal processes for the removal of content. We are further concerned that authoritarian countries often take inspiration from these kinds of requirements to put more direct pressure on companies to censor content.

**Recommendation:**

- Any legislation in this area should eschew powers to i) disrupt business activities, ii) undertake ISP blocking, or iii) implement a regime for senior management liability.

**Question 14 - Appeals and judicial review**

39. **ARTICLE 19 believes that there should be** a statutory mechanism for companies to appeal against decisions of the regulator. Companies should be able to challenge decisions to impose fines or website blocking. Equally, senior management should be able to challenge decisions made against them. Any appeal should be reviewed on the merits of the case rather than on judicial review principles given the significance of the interference with users' right to freedom of expression.

**Recommendation:**

- Decisions of the regulator should be subject to appeal on the merits.